



USAID
DEL PUEBLO DE LOS ESTADOS
UNIDOS DE AMÉRICA

**REFORMA
EDUCATIVA
EN EL AULA**

PROGRESS PROFILES OF STUDENTS AND TEACHERS: COMBINING GROWTH MODELS TO DESCRIBE LEARNING ANALYSIS WHIP



MARCH 2014

This material was developed under Work Order No. EDH-I-00-05-00033 of the U.S. Agency for International Development, Guatemala Mission (USAID/G), with Juarez and Associates: USAID/Education Reform in the Classroom project, and in support of the Grant Agreement for Strategic Objective No. 520-0436.7, "Social Investment: Healthier persons with a better education"

USAID/ EDUCATION REFORM IN THE CLASSROOM

**PROGRESS PROFILES OF STUDENTS AND TEACHERS:
COMBINING GROWTH MODELS TO DESCRIBE
LEARNING**

**ANALYSIS OF THE WESTERN HIGHLANDS INTEGRATED
PROGRAM - WHIP**

Prepared for

United States Agency for International Development, Guatemala

Contractor

Juárez and Asociados, Inc.

Prepared by

Leslie Rosales

Guatemala, March 2014

**Contrato No. EDH-I-00-05-00033-00 y
Orden de Trabajo EDH-I-05-05-00033-00**

Table of Contents

EXECUTIVE SUMMARY	6
Methodology	7
Methodology for second, third and fourth grades.....	7
Methodology for first grade.....	8
Results	9
First grade profile	9
Second grade teachers' progress profile	9
Third grade progress profile	10
Fourth grade progress profile.....	10
Conclusions	11
Implications.....	13
I. INTRODUCTION.....	14
II. METHODOLOGY.....	15
Second, third and fourth grades.....	15
Does a student read more at the end of the year than he did at the beginning of the year? (Growth relative to Self)	16
Is a student's reading progressing in relation to his peers from the beginning of the year to the end? (Growth relative to others).....	17
Is a student's reading closer to the established reading standard for the grade at the end of the year than his reading was at the beginning of the year? (Growth relative to standard)	17
What type of items can students answer at the beginning of the year and, what type of items can they answer at the end of the year?	18
III. ILLUSTRATIONS OF PROFILES OF STUDENTS AND GROUPS.....	18
IV. RESULTS.....	22
First grade	22
Second Grade.....	27
Third grade	32
Fourth Grade	37
V. CONCLUSIONS.....	41
VI. IMPLICATIONS.....	43
VII. REFERENCES	45

EXECUTIVE SUMMARY

The main purpose of this study was to answer to the question: How does the reading level of children improve during the school year in the five departments where the Western Highlands Integrated Program - WHIP works? Implicit in this research question is the notion of growth or progress in reading.

This study combined four models that address different perspectives of reading growth, which allowed the creation of progress profiles. A progress profile, in this study, is a story that provides a concrete example of growth in reading within a year for the different grades, and takes into account different conditions such as sex, language, grade repetition, preschool, teacher, among others. The ultimate goal of this study is to outline what does it means to grow in reading for the students that participated in the WHIP program.

According to Castellano and Ho, contrary to status, growth or progress “describes the academic performance of a student or group over two or more time points (2013, p. 13). Status, however, describes performance at a single point in time (Castellano & Ho, 2013, p. 12). In other words, to understand progress of student learning, beyond considering current achievement, one must take into account prior achievement (Betebenner, 2009, p. 43). However, it is important to clarify, that taking into account prior achievement does not imply attributing causality of current achievement to prior achievement; instead prior achievement is taken as a starting point to progress.

Growth model literature is extensive (Ho & Castellano, 2013). Most models have been developed in the United States to answer accountability questions related to the No Child Left Behind Policy (Linn, Baker, & Betebenner, 2002). The models were built to demonstrate progress or growth based on different educational accountability questions; therefore, their focus, statistical foundations and interpretation of results vary greatly. The decision to adopt any of these models to answer a growth question depends on factors related to: 1) the research question, 2) the data available, and 3) the audience interested in results.

From the above, it is important to stress that the purpose of this study was to understand how does the reading level of children improve during the school year in the five departments where the Western Highlands Integrated Program – WHIP works? This question denotes a descriptive study where one can demonstrate, under different perspectives, improvement in reading. Thus, in this study growth was described by combining four models of growth: (a) Growth relative to self or the gains model (DePascale, 2006), (b) Growth relative to others (D. W. Betebenner, 2008), (c) Growth relative to standards (D. Betebenner, 2009) and, (d) Growth relative to items. The combination of these four models allowed the creation of progress profiles at the student and group levels.

Methodology

As mentioned before, there are different approaches to growth that one could take to answer to the research question of this study: how does the reading level of children improve during the school year in the five departments where the Western Highlands Integrated Program – WHIP works? In this section, a brief description of each approach and the model corresponding to it are provided.

The data collection was done in 100 public schools distributed in the five departments of the WHIP. The collection was done in two points in time, at the beginning and at the end of year 2013. The beginning of the year collection was done in February and March and end of the year collection was done in August and September of the same year; therefore, there is a six-month to an eight-month span between the measures. Tests were given in a Mayan mother language and Spanish as a second language.

Each approach of growth: (a) growth or progress relative to self, (b) growth or progress relative to others, (c) growth or progress relative to a standard and (d) growth or progress relative to the items in the test, was analyzed with a different methodology. For grades second, third, and fourth this methodology is consistent. However, for first grade the methodology varies because a different test was given to students at the two time points.

Methodology for second, third and fourth grades

Does a student read more at the end of the year than he did at the beginning of the year? (Growth relative to Self)

To know if second, third and fourth grade children have made progress within a school year, a gains model is proposed (DePascale, 2006). The gains model supports progress description. The model gives an answer to how much a student learned on an absolute scale; therefore, scores must be on a common scale to be meaningful. However, in the case of second, third and fourth grades of this study this is not a concern because the same test was given at the beginning and the end of the year.

Gain scores are estimated by subtracting current status minus initial status. In this study, for grades second, third and fourth, gain scores will be estimated by subtracting end of the year theta scores¹ minus beginning of the year theta scores. Alternatively, end of the year raw scores minus beginning of the year raw scores were subtracted. Furthermore, the sign of the gain score has a meaning, positive gains represent improvement and negative gains a decline. Finally, an average of gain scores allows group-level interpretations of progress.

¹ Measure of ability using Item Response Theory.

Is a student's reading progressing in relation to his peers from the beginning of the year to the end? (Growth relative to others)

Additionally for responding if a student has made progress in reading, one might want to compare his progress to his peers; the question then is if a student is progressing from one point to the other better, worse or the same as students of the same cohort. For this purpose, the proposed model is the Student Growth Percentile (SGP) (D. Betebenner, 2009). In a few words, the SGP describes the location of a student's current score in relation to the current scores of students with similar score histories or similar scores at the beginning of the year.

Finally, a median SGP can serve for group-level analyses. In this study, group level reports will be provided by sex, language, grade repetition, and preschool education.

Is a student's reading closer to the established reading standard for the grade at the end of the year than his reading was at the beginning of the year? (Growth relative to standard)

Another measure of progress is to explain if a student achieved the standard or is closer to achieving it at the end of the school year than he was at the beginning of the year. In order to answer to this question, the difference between the student theta and the theta cut point for the grade will be used to estimate the proximity to achieving the standard in both points in time. In the case of second, third and fourth grades the interpretation is transparent as the same test was used at the beginning and end of the year. That is, the closer to 0 is the difference between scores and the standard, the more progress the student has made. Average proximity to the standard for the two points in time will be used for group-level interpretations. Alternatively, the number of students that achieved grade standard at the end of the year compared to those at the beginning of the year was also computed for this growth measure.

What type of items can students answer at the beginning of the year and, what type of items can they answer at the end of the year?

Learning to read is a process. Therefore, beyond estimating progress or growth under different statistical models, it is also important in this study to illustrate the continuum of acquiring the ability to read. Along with it, it is a purpose of this study to demonstrate the types of items that students can read at the beginning of the year and the end of the year.

This will be done by estimating the "average gain score" (DePascale, 2006) at the beginning of the year for the grade and the aggregations proposed for this study: sex, language, repetition and preschool. The same way the "average gain score" will be estimated for the end of the year scores. This will allow demonstrating the types of items that a student can do when progression.

Methodology for first grade

First grade students, as we have mentioned before, took two different tests at the two time points of this study: (a) LEE² and (b) National Reading Assessment for first

² Emergent reading test

grade. Despite the fact that the two tests assess the same construct, reading, they vary greatly. Having made that clear, the methodology that was described previously for second, third, and fourth grades is not suitable for first grade. Consequently, a methodology to describe progress in first grade will consist in relating the starting point of first grade students with their ability at the end of the year. That is, the first task into describing growth is to confirm that there is a relationship between the two tests. This was not a concern in the other three grades because students took the same test.

To explore the relationship between the tests, LEE and the National Assessment binary logistic regression was used. Chi square analyses were performed to confirm relationships between the two tests. In the procedures, the predictors were: a) the letter identification subtest of LEE, b) the emergent writing subtest of LEE, and c) the reading subtest of LEE. The dependent variable in the logistic regression was being classified in the “satisfactory” level of achievement in the end of the year national test for first grade (level 3, theta score= 0.201).

More specifically, students were classified into different starting criteria based on LEE, the criteria was then related to the end of the year achievement obtained of the National Assessment of first grade. That way, one would be able to answer the following question: How many students from each criterion in LEE achieved the standard at the end of the year?

Furthermore, progress profiles in first grade could also be created. However, these profiles are focused in achieving first grade standard. In the following section a detailed description of the results is provided for each grade.

Results

First grade profile

*Progress estimations of **First** grade students that participated in the WHIP program were done with a valid sample of 1616 students. From the sample 48% were girls and 52% were boys; 26% reported speaking a Mayan Language and the rest reported speaking Spanish. 31% did not attend preschool and the rest reported attending. Finally, 33% of the students had repeated a grade at least once. Students in first grade who know the alphabet principle more, and who write and read more words phonetically achieve the standard more at the end of the grade. Those students in first grade who achieved the standard can read complete sentences that match an illustration.*

Second grade teachers' progress profile

*Progress estimations of **Second** grade teachers that participated in the WHIP program were done with a valid sample of 90 teachers. From the sample 62.8% were female teachers and 37.2% were male teachers. 54.3% had less than 9 years of experience and the rest had at least 9 years of teaching experience; 9 years was the average years of experience. Students that declined were clustered in both male and*

female teachers, as well as in those who had more than the mean years of experience or less than the mean years of experience (9 years). Those teachers with more experience than the average, had students that progress more in relation to students that started with the same ability (median SGP= 50) than those teachers with less experience than the average (median SGP= 46).

Third grade progress profile

Progress estimations of **Third** grade students that participated in the WHIP program were done with a valid sample of 1668 students. From the sample 45.9% were girls and 47.5% were boys; 42.1% were ladino students, 50.5% were Mayan, 7 students were Garífuna and 5 students were Xinka. Finally, 24.7% of the students had repeated a grade at least once. At the beginning of the year third grade students could only answer 12 items correctly in the national reading assessment for third grade in Spanish. However, by the end of the year they answered 14 correctly on average, which represents a gain of 2 points. On average, their ability improved 0.21. An example of the type of items they can answer now is finding the details in a five-sentence paragraph. This item is harder than they could do at the beginning of the year where they could only find the main character in a four-sentence paragraph. Unfortunately, about 37.9% declined in their ability by the end of the year. By the end of third grade, on average students still had not achieved third grade reading standard but they are closer to achieving it than they were at the beginning of third grade. Furthermore, the number of students in the “satisfactory” level increased at the end of the year, and therefore, the number of students in the lower levels declined. Girls progressed more in relation to their peers that started with the same ability.

Fourth grade progress profile

Progress estimations of **Fourth** grade students that participated in the WHIP program were done with a valid sample of 1605 students. From the sample, 49% were girls and 51% were boys; 38.6% were ladino students, 54.5% were Mayan, 3 students were Garífuna. Finally, 24.7% of the students had repeated a grade at least once. At the beginning of the year fourth grade students could only answer 12 items correctly in the national reading assessment for third grade in Spanish. However, by the end of the year they answered 14 correctly on average, which represents a gain of 2 points. On average, their ability improved 0.21. An example of the type of items they can answer now is inferring the purpose of the author in a four-sentence story. This item is harder than they could do at the beginning of the year where they could only identify details of a short paragraph. Unfortunately, about 37.9% declined in their ability by the end of the year. Since the beginning of fourth grade, on average students had achieved third grade reading standard but they did not make it on average to excel third grade standard at the end of fourth grade. Girls progressed more in relation to their peers that started with the same ability.

Conclusions

All the models used in this study coincide in results to describe improvement in reading. This study allowed describing improvement under different perspectives of growth. Table 9 summarizes the different estimations done for each of the approaches to growth in each grade.

Table 9: Summary statistics of growth under each approach

		GRADE AVERAGE	BOYS	GIRLS	REPEATERS	NON REPEATERS
Growth relative to self: GAIN SCORE (THETA SCALE)**	SECOND	0.45	0.43	0.46	0.44	0.46
	THIRD	0.21	0.19	0.22	0.25	0.18
	FOURTH	0.21	0.21	0.22	0.22	0.2
Growth relative to self: GAIN SCORE (ABSOLUTE SCALE)**	SECOND	2	2	3	2	2
	THIRD	2	1	2	2	2
	FOURTH	2	1	2	2	2
Growth relative to others: SGP*	SECOND	50	48	51	43	54
	THIRD	49	45	53	49	49
	FOURTH	49	47	52	49	50
Growth relative to a standard: Percentage of standard achievers, at the end of the year	SECOND	52%	50%	54%	42%	58%
	THIRD	40%	38%	42%	37%	42%
	FOURTH	64%	63%	65%	64%	64%

Children that participated in the WHIP program improved in reading; however the improvement is low. Overall, children in second, third and fourth grades improved during the school year in the five departments where the Western Highlands Integrated Program – WHIP. Gains ranged from 1 to 3 points on the absolute scales, and between 0.18 and 0.46 on the theta scale.

Students improved more in second grade than in the upper grades. Students in second grade gained between 0.43 and 0.46 points in the theta scale; however, in the upper grades (third and fourth) the gained was between 0.2 and 0.22 points.

Children that participated in the WHIP program improved less in reading in the Mayan languages than they do in Spanish. Students showed a little improvement in K'iche'. And, with the exception of grade three, there was no improvement in second and fourth grades (See Table 10).

Table 10: Absolute gain in each language by grade

	Second	Third	Fourth
Gain in the absolute scale K'iche'	2	1	No data available
Gain in the absolute scale Mam	0	1	0

A high percentage of students that participated in the WHIP program declined in ability. Most students gained ability; however, in second grade 33.7% of students declined. In third the percentage was 37.9 and in fourth grade 32.8.

Girls that participated in the WHIP program progressed more in reading than boys. In all the models used to describe progress in this study, girls had higher estimations than boys (See Table 9).

Repeaters improved less than their non-repeaters peers. Non-repeaters and repeaters had about the same gains during the year, except in third grade, where repeaters gained more. However, in relation to their peers, repeaters consistently progressed less than their non-repeater peers. Furthermore, non-repeaters achieved the standard more than repeaters.

Students that know the alphabetic principle, read words and write some words, are more likely to achieve first grade standard at the end of the year. The criteria to classify students in LEE are positively related to the level of achievement at the end of first grade. First grades with higher scores in the Emergent reading assessment LEE at the beginning of the year are more likely to achieve first grade standard at the end of the year. Similarly, those students with lower scores in LEE are more likely to be classified in the “unsatisfactory” and “needs improvement” levels of achievement.

Students that participated in the WHIP program are a year behind in terms of achieving the standard. Students in first grade on average do not achieve first grade standard at the end of first grade or second grade. However, they are closer to achieving it at the end of second grade. Third graders do not achieve third grade standard at the end of the year but they do in fourth grade. Although, on average fourth grade students achieve third grade standard, they do not excel it at the end of the year.

It is possible to build a learning continuum from the average items that students read at each grade and time point. In terms of the skills that students have, it was possible to build a progress latter that illustrates the skills that students have at each grade from the typical items of entry and exit of each grade. Figure illustrates such progress in terms of skills students have.

Table 11: Illustration of progress in skills

End of the year →	LEE	Fourth grade	
		Infers author's purpose in a paragraph.	Remembers details of a four-sentence paragraph.
		Third grade	Remembers details of a three-sentence paragraph
		Identifies the main character of a three-sentence paragraph	
Beginning of the year →		Second grade	
		Finds the appropriate word to complete a paragraph.	Read two words to match a picture (one word is a name).
		First grade	Read one word that matches a picture.
		Reads words	Knows alphabet

Implications

Measuring learning implies longitudinal studies. As with this study, measuring learning implies collecting achievement information at least twice for each student. This methodology implies complex logistics; however, conclusions about learning are more robust.

Interventions focused to teachers should be based on learning and progress information. In this study it was possible to create profiles of teachers focused on their students' learning or progress. This will allow creating interventions that are focused to teachers that need more help in bringing their students to a higher achievement level.

Progress information should be based on different models that coincide in results. Descriptions or conclusions about progress or learning should be based in different models that coincide in results. This methodology allows stronger conclusions at the individual and group levels.

An effort to communicate progress/growth in reading to teachers should be made. Illustrations of a learning continuum in each grade should be provided to teachers that orient their teaching practice.

I. INTRODUCTION

The main purpose of this study is to answer to the question: How does the reading level of children improve during the school year in the five departments where the Western Highlands Integrated Program - WHIP works? Implicit in this research question is the notion of growth or progress in reading.

This study combined four models that address different perspectives of reading growth, which allowed the creation of progress profiles. A progress profile, in this study, is a story that provides a concrete example of growth in reading within a year for the different grades, and takes into account different conditions such as sex, language, grade repetition, preschool, teacher, among others. The ultimate goal of this study is to outline what does it mean to grow in reading for the students that participated in the WHIP program.

According to Castellano and Ho, contrary to status, growth or progress “describes the academic performance of a student or group over two or more time points (2013, p. 13). Status, however, describes performance at a single point in time (Castellano & Ho, 2013, p. 12). In other words, to understand progress of student learning, beyond considering current achievement, one must take into account prior achievement (Betebenner, 2009, p. 43). However, it is important to clarify, that taking into account prior achievement does not imply attributing causality of current achievement to prior achievement; instead prior achievement is taken as a starting point to progress.

Growth model literature is extensive (Ho & Castellano, 2013). Most models have been developed in the United States to answer accountability questions related to the No Child Left Behind Policy (Linn, Baker, & Betebenner, 2002). The models were built to demonstrate progress or growth based on different educational accountability questions; therefore, their focus, statistical foundations and interpretation of results vary greatly. The decision to adopt any of these models to answer a growth question depends on factors related to: 1) the research question, 2) the data available, and 3) the audience interested in results.

From the above, it is important to stress that the purpose of this study is to understand how does the reading level of children improve during the school year in the five departments where the Western Highlands Integrated Program – WHIP works³? This question denotes a descriptive study where one can demonstrate, under different perspectives, improvement in reading. Thus, different models are proposed in this study to describe learning at the student and group-level.

There are at least four approaches under which learning to read in the schools where WHIP works could be described:

1. Growth relative to self
2. Growth relative to others
3. Growth relative to standards
4. Growth relative to items

³ Huehuetenango, Quetzaltenango, Quiché, San Marcos and Totonicapán.

The first one refers to whether a student reads more at the end of the year than he or she did at the beginning of the year. This was defined by DePascale (2006) as “growth relative to self” or the gains model, in which the basic idea is to obtain the difference between the measures at two points in time, the end of the year minus the beginning of the year.

The second one refers to how a student’s reading is progressing in relation to his peers from the beginning of the year to the end. In which case, improvement is interpreted in a normative way. This was defined by DePascale (2006) as “growth relative to others”. Currently, a normative model greatly used is the student Growth Percentile Model - SGP (D. W. Betebenner, 2008).

Furthermore, one might be interested in knowing whether a student’s reading is closer to the established reading standard for the grade at the end of the year than his reading was at the beginning of the year. This approach is known as “growth relative to a standard” (D. Betebenner, 2009; DePascale, 2006). The basic idea of this model is to obtain the difference between student’s achievement and the grade standard at two points in time. The difference should be the same, or should be above the standard at the end of the year; however, it could be the case that is less for those students who don’t achieve the standard by the end of the year.

Lastly, a relevant question could be, what does it mean to progress in reading from the beginning of the year to the end of the year. What types of items can a student respond correctly at the end of the year that he could not at the beginning of the year as a product of instruction? This approach requires a qualitative analysis of the items placed in the typical growth profiles of the different cohorts.

In the following section a detailed description of the methodology proposed is provided. The data for first grade had to be treated differently, because students did not take the same test at the beginning and the end of the year and both tests differ substantially, despite the fact that they assess the same construct.

II. METHODOLOGY

Second, third and fourth grades

As mentioned before, there are different approaches to growth that one could take to answer to the research question of this study: how does the reading level of children improve during the school year in the five departments where the Western Highlands Integrated Program – WHIP works? In this section, a detailed description of each approach and the model corresponding to it are provided.

The data collection was done in 100 public schools distributed in the five departments of the WHIP. The collection was done in two points in time, at the beginning and at the end of year 2013. The beginning of the year collection was done in February and March and end of the year collection was done in August and September of the

same year; therefore, there is a six-month to an eight-month span between the measures. Tests were given in a Mayan mother language and Spanish as a second language. Table 1 displays the tests of the data collection that was used for the analysis.

Table 1: Tests data collected at the beginning and end of the year

Grade	Beginning of the year measure	End of the year measure
First	Emergent Reading Test (LEE)	National Reading Test for first grade (Spanish, Mam and K'iche')
Second	National Reading Test for first grade (Spanish, Mam and K'iche')	National Reading Test for first grade (Spanish, Mam and K'iche')
Third and Fourth	National Reading Test for third grade (Spanish, Mam and K'iche')	National Reading Test for third grade (Spanish, Mam and K'iche')

Each approach of growth: (a) growth or progress relative to self, (b) growth or progress relative to others, (c) growth or progress relative to a standard and (d) growth or progress relative to the items in the test, was analyzed with a different methodology. For grades second, third, and fourth this methodology is consistent. However, for first grade the methodology varies because a different test was given to students at the two time points. The details of the methodology used for first grade will be described in a separate section.

Does a student read more at the end of the year than he did at the beginning of the year? (Growth relative to Self)

To know if second, third and fourth grade children have made progress within a school year, a gains model is proposed (DePascale, 2006). The gains model supports progress description. The model gives an answer to how much a student learned on an absolute scale; therefore, scores must be on a common scale to be meaningful. However, in the case of second, third and fourth grades of this study this is not a concern because the same test was given at the beginning and the end of the year.

Gain scores are estimated by subtracting current status minus initial status. In this study, for grades second, third and fourth, gain scores will be estimated by subtracting end of the year theta scores⁴ minus beginning of the year theta scores. Alternatively, we will subtract end of the year raw scores minus beginning of the year raw scores. Finally, the sign of the gain score has a meaning, positive gains represent improvement and negative gains a decline.

Furthermore, an average of gain scores allows group-level interpretations of progress. In this study, group level reports will be performed by sex, language, grade repetition, and preschool education. However, it is important to keep in mind that the

⁴ Measure of ability using Item Response Theory.

distribution of the gain scores matters when interpreting group-level progress: “A near zero average gain score indicates that either all students had near zero gains or that there was rough balance between positive gains and negative gains that average to near zero. A positive average gain score indicates that students, on average, made positive gains, whereas a negative average gain score indicates that students generally declined in performance” (Ho & Castellano, 2013, p. 39). In other words, the same gain score can represent different scenarios of learning in a group. For example, a group of student might get an average gain score of two, but have some students that decline and others that improve. Contrary, a group of students could have a gain score of two and have all the students in that group with positive gains. That is why simple summary statistics are not sufficient for this study; an analysis of variance will be done at the two points in time. Ideally, variance at the end of the year should decrease if instruction was effective.

Is a student’s reading progressing in relation to his peers from the beginning of the year to the end? (Growth relative to others)

Additionally for responding if a student has made progress in reading, one might want to compare his progress to his peers; the question then is if a student is progressing from one point to the other better, worse or the same as students of the same cohort. For this purpose, the proposed model is the Student Growth Percentile (SGP) (D. Betebenner, 2009). In a few words, the SGP describes the location of a student’s current score in relation to the current scores of students with similar score histories or similar scores at the beginning of the year. For example, a student earning an SGP of 80 progressed as well or better than 80% of his peers of the reference group. One shortcoming of the SGP model is that it requires large number of students for the SGP estimation to be stable; in other words if only a few students have similar scores at the beginning of the year the rank estimation becomes unstable. However, data fit analysis are available for the model that will be provided in this analysis. For illustration purposes the basic formula to estimate a percentile rank in the SGP model is the following (Ho & Castellano, 2013, p. 94):

$$\text{percentile rank} = \frac{\text{Number of students at or below a particular score} + (.5 * \text{Number of students at score})}{\text{Number of students in the academic peer group}} \times 100$$

However, the actual SGP does not divide the cohort in students with the exact starting point score as this formula shows because the groups become smaller. Instead the SGP uses quantile regression. Finally, a median SGP can serve for group-level analyses. In this study, group level reports will be performed by sex, language, grade repetition, and preschool education.

Is a student’s reading closer to the established reading standard for the grade at the end of the year than his reading was at the beginning of the year? (Growth relative to standard)

Another measure of progress is to explain if a student achieved the standard or is closer to achieving it at the end of the school year than he was at the beginning of the year. In order to answer to this question, the difference between the student theta and the theta cut point for the grade will be used to estimate the proximity to achieving the

standard in both points in time. In the case of second, third and fourth grades the interpretation is transparent as the same test was used at the beginning and end of the year. That is, the closer to 0 is the difference between scores and the standard, the more progress the student has made. Average proximity to the standard for the two points in time will be used for group-level interpretations.

What type of items can students answer at the beginning of the year and, what type of items can they answer at the end of the year?

Learning to read is a process. Therefore, beyond estimating progress or growth under different statistical models, it is also important in this study to illustrate the continuum of acquiring the ability to read. Along with it, it is a purpose of this study to demonstrate the types of items that students can read at the beginning of the year and the end of the year.

This will be done by estimating the “average gain score” (DePascale, 2006) at the beginning of the year for the grade and the aggregations proposed for this study: sex, language, repetition and preschool. The same way the “average gain score” will be estimated for the end of the year scores. This will allow demonstrating the types of items that a student can do when progression.

III. ILLUSTRATIONS OF PROFILES OF STUDENTS AND GROUPS

As we mentioned before, looking at growth under different perspectives, allows the creation of progress profiles at the student and group levels. In this section different progress profiles are presented based on the information obtained per student under the different models. This information is detailed in Tables 2 and 3. Furthermore, group level interpretations are also exemplified based on the same information.

Finally, the data for this study, as well as the one obtained with the models, allowed the creation of teacher progress profiles, in which it is possible to describe the reading progress of his or her students for the year 2013.

Student growing profiles

The following are two examples of student progress profiles that could be created based on the information obtained from the different growth models described in the previous section.

María is a second grade student who turned 10 in 2013. She is Mayan and she is in Miss [206] classroom of school [09-13-0375-43]. At the beginning of the year she could only answer 9 items correctly in the national reading assessment for first grade in Spanish. However, by the end of the year she answered 13 correctly, which represents a gain of 4 points, and a 0.924 progress in the ability scale. An example of the type of items she can answer now is finding the main character in a three-sentence story. This item is harder than she

could do at the beginning of the year where she could only read single-sentence items. By the end of second grade, she still has not achieved first grade reading standard but she is closer to achieve it than she was at the beginning of second grade. In relation to her peers with the similar ability at the beginning of second grade, she has made substantial progress; she performed better than 71% of her peers.

Table 2: Progress information for student 574

ID	574
Departamento	9
Municipio	13
School	09-13-0375-43
Grade	2
Student Name	Maria
Age at the beginning of the year	9
Age at the end of the year	10
Ethnicity	Maya
Teacher	206
Achievement at the beginning of the year	2
Achievement at the end of the year	3
Ability at the beginning of the year	-0.406
Ability at the end of the year	0.518
Absolute score at the beginning of the year	9
Absolute score at the end of the year	13
Gain in ability	0.924
Gain in absolute score	4
SGP	71
Ability difference with cut point at the beginning of year	1.67
Ability difference with cut point at the end of year	0.74

Nelson is a second grade student who turned 10 in 2013. He is ladino and he is in Miss [309] classroom of school [12-24-0963-43]. At the beginning of the year he could only answer 3 items correctly in the national reading assessment of first grade in Spanish. However, by the end of the year he answered 16 correctly, which represents a gain of 13 points, and a 2.81 progress in the ability scale. An example of the types of items he is able to answer now is a prediction item, where he chooses the sentence that best continues a short story. This is harder than he could do at the beginning of the year, where he could only read single-word items. By the end of second grade, he achieved first grade reading standard. In relation to his peers with the similar ability at the beginning of second grade, Nelson has made great progress; he performed better than 95% of his peers.

Table 3: Progress information for student 37

ID	37
Departamento	12
Municipio	24
School	12-24-0963-43
Grade	2
Student Name	Nelson
Age at the beginning of the year	9
Age at the end of the year	10
Ethnicity	Ladino
Teacher	309
Achievement at the beginning of the year	2
Achievement at the end of the year	3
Ability at the beginning of the year	-1.54
Ability at the end of the year	1.263
Absolute score at the beginning of the year	3
Absolute score at the end of the year	16
Gain in ability	2.812
Gain in absolute score	13
SGP	95
Ability difference with cut point at the beginning of year	2.81
Ability difference with cut point at the end of year	-0.00201

Group-level growing profiles

The following is an illustration of a group-level progress profile in learning to read in Spanish that could be created for second grade based on the information obtained from the different growth models described in the previous section.

On average, students in second grade in the Western Highlands Integrated Program - WHIP gain about 2 points (0.44 on ability scale) on first grade test in second grade from the beginning to the end of the year. This pattern was similar for boys and girls in second grade. There were a number of students, regardless of their sex, whose scores decline instead of improving. Most boys gain 1 point in the test at the end of the year and the girls gain 3 points.

Teacher profiles

The following is an illustration of a teacher progress profile that could be created based on the information obtained from the different growth models described in the previous section and the data available for the WHIP program.

Miss [252]

Miss [252] is a teacher in Cuilco, Huehuetenango. She teaches in second grade, she has a group of 20 students. At the beginning of the year, her students could answer, on average, 7 questions correctly in the national reading assessment for first grade. Such questions were mostly single sentence items (average theta of $-.4654$). However, by the end of second grade her students were able to answer, on average, 15 questions correctly as well as harder questions (average theta of $=.97525$), such as identifying the main idea in a

short paragraph. All of her students showed gain in their scores by the end of the year, the student who gained less, gained 1 point and the one who gained the most gained 11 points. Miss [252] students performed better than 87% of the students in WHIP program with similar ability at the beginning of the year. Finally, Miss [252] students achieved first grade standard by the end of second grade.

Miss [14]

Miss [14] is a second grade teacher in San Juan Cotzal, Quiché. She leads 11 students in her classroom. At the beginning of the year, her students answered, on the average, 8 questions correctly on the Spanish national reading assessment for first grade. At the beginning of the year students could read single sentence items (average theta of $=-.8139$). Unfortunately, at the end of second grade, students in Miss [14] grade declined and could only answer, on average, 3 questions correctly; these questions required reading a single word (average theta of $=-.2.04$). Miss [14] students performed better than 6% of the students in WHIP program with similar ability at the beginning of the year. Finally, Miss [14] students did not achieve first grade standard by the end of second grade.

First grade

First grade students, as we have mentioned before, took two different tests at the two time points of this study: (a) LEE⁵ and (b) National Reading Assessment for first grade. Despite the fact that the two tests assess the same construct, reading, they vary greatly. Having made that clear, the methodology that was described previously for second, third, and fourth grades is not suitable for first grade. Consequently, a methodology to describe progress in first grade will consist in relating the starting point of first grade students with their ability at the end of the year. That is, the first task into describing growth is to confirm that there is a relationship between the two tests. This was not a concern in the other three grades because students took the same test.

To explore the relationship between the tests, LEE and the National Assessment binary logistic regression was used. Chi square analyses were performed to confirm relationships between the two tests. In the procedures, the predictors were: a) the letter identification subtest of LEE, b) the emergent writing subtest of LEE, and c) the reading subtest of LEE. The dependent variable in the logistic regression was being classified in the “satisfactory” level of achievement in the end of the year national test for first grade (level 3, theta score= 0.201).

More specifically, students were classified into different starting criteria based on LEE, the criteria was then related to the end of the year achievement obtained of the National Assessment of first grade. That way, one would be able to answer the following question: How many students from each criterion in LEE achieved the standard at the end of the year?

The following table describes the starting point criteria of each subtest in LEE.

⁵ Emergent reading test

Table 4: Criteria to score LEE in first grade

	Criteria	Description of ability
Letter identification	<ul style="list-style-type: none">- Up to 6 points- 7 to 12 points- 13 to 20 points- 21 points or higher	Students, who earned more points, could name letters without a prompt or help.
Emergent writing	<ul style="list-style-type: none">- Up to 2 points- 3 to 7 points- 8 to 12 points- 12 points or higher	Students, who earned more points, could write words phonetically, either words of their choice or dictated by the data collector.
Emergent reading	<ul style="list-style-type: none">- No points- 1 to 2 points- 3 to 5- 6 points or higher	Students, who earned more points, could read more words.

Furthermore, progress profiles in first grade could also be created. However, these profiles are focused in achieving first grade standard. In the following section a detailed description of the results is provided for each grade.

IV. RESULTS

First grade

As has been described before the data for first grade had to be treated differently, because students did not take the same test at the beginning and the end of the year and both tests differ substantially, despite the fact that they assess the same construct. In the following section, results of the relationship between the two tests and patterns of achievement depending on the students' score in LEE will be presented.

Results of binary logistic regression

Overall the criteria of LEE significantly predicts achieving the standard at the end of the year ($\chi^2 = 112.07$, $df=3$), $p < 0.05$. The model allowed identifying 21.3% those who achieve and 93.5% those who did not achieve. The odds of achieving the standard at the end of the year are greater to those with higher scores in each criterion in LEE than for those with lower scores.

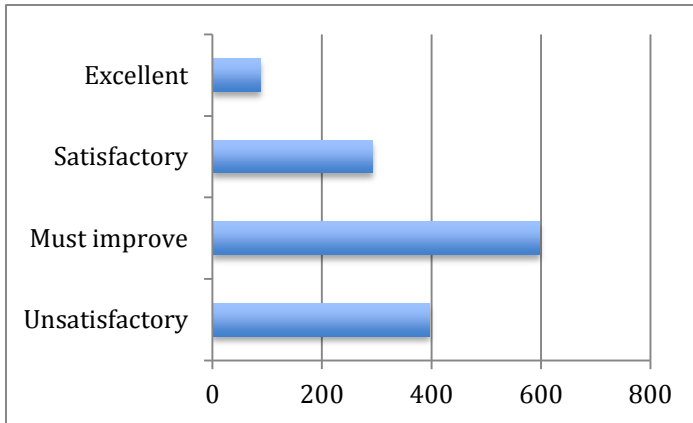


Figure 1: Number of first grade students in each achievement level at the end of the year

Moreover, after controlling for scores of the emergent writing subtest, and reading, the likelihood of achieving the standard is higher for those who identified more letters without help in the letter identification subtest, $X^2_{Wald}(1, N = 1616), p < 0.05$, Odds ratio = 1.523. Similarly, after controlling for the writing scores and the scores in letter identification subtest, the likelihood of achieving the standard at the end of the year is higher for those who could read more words the reading subtest, $X^2_{Wald}(1, N = 1616), p < 0.05$, OR = 1.373.

However, after controlling for scores in the letter identification, and scores in reading, there is no significant difference in the likelihood of achieving the standard for those who write better in the writing test, $X^2_{Wald}(1, N = 1616), p = 0.217$, OR = 1.10. An explanation to this result is that writing might be a mechanic process of reproducing or copying teacher's writing that does not require student's effort to sound out letters.

The relationship between the criteria in LEE and achieving the standard at the end of the year was confirmed with Chi square statistics. The relationship between the criteria in LEE and achieving the standard at the end of the year was confirmed by sex, repetition, and language. Although the logistic regression did not identified the emergent writing as a significant predictor after controlling for others, it is possible to see the same pattern of achievement as with the other predictors, that is those who earned more points in this test, were classified more into the satisfactory and excellent levels of achievement (See Figure 4). This pattern was true for those who attended preschool ($\chi^2 = 58.01, N = 940, p < 0.05$) but not for those who did not ($\chi^2 = 14.97, N = 426, p = 0.92$). This suggests a possible interaction between attending preschool and achieving the standard, depending on the starting criteria of LEE (See Table 5).

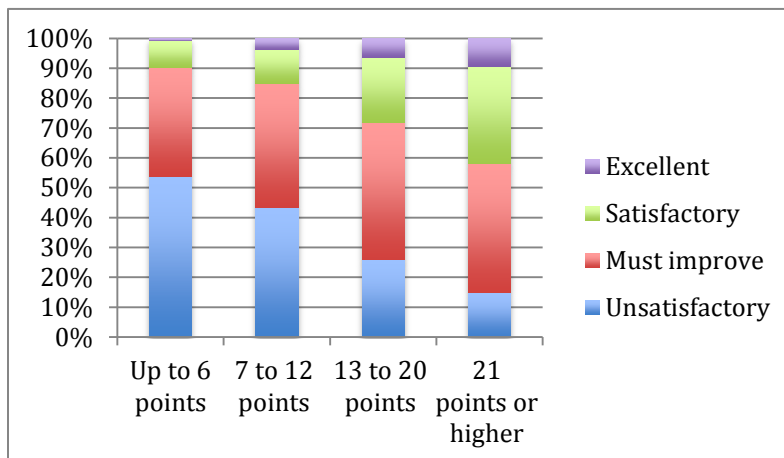
Table 5: Chi square statistics for first grade

	N	Letter recognition Pearson Chi-Square	Writing	Reading
Preschool				
Did not attend	426	34.995**	14.968	21.148**
Attended	940	86.752**	58.018**	87.176**
Sex				
Girl	655	60.618**	44.563**	65.345**
Boy	720	79.739**	33.396**	56.576**
Repetition				
Did not repeat	906	98.030**	68.323**	94.069**
Repeated	451	46.560**	26.433**	39.165**
Language				
Mayan	352	18.222**	24.435**	43.257**
Spanish	1002	103.094**	58.677**	74.022**

** p<0.001

How many students from each criterion in LEE achieved the standard at the end of the year?

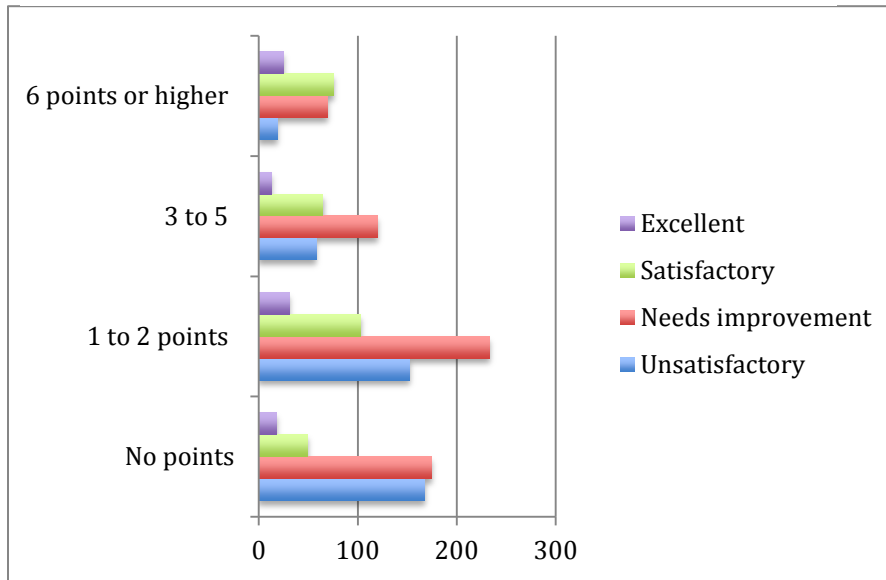
Figure 2: Percentage of students classified in each achievement level from each criteria in letter identification of LEE



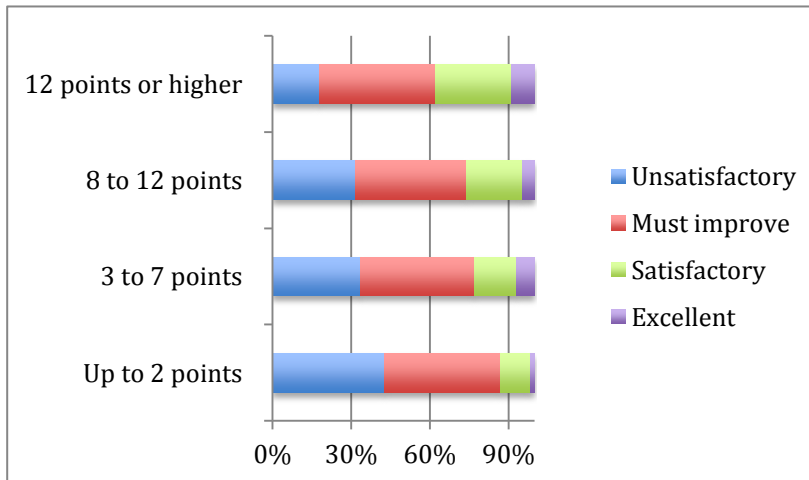
About 21% of students were classified into the “satisfactory” level at the end of first grade; 37.9% were classified in the “must improve” level; 25.2% were classified “unsatisfactory” and only 5.5% were classified as excellent (N= 1375). That is, about 28% of students achieved the standard by the end of the year. As we learned from the logistic

regression, the letter identification and the reading subtests were significant predictors of achievement, in Figure 2, it is possible to see the obvious relationship that the more letters students could identify the more the students could achieve the satisfactory or excellent, levels of achievement. This pattern was true for those who attended preschool and those who did not. The pattern was also true for boys and girls, repeaters and non-repeaters and for Mayan students and ladino students.

Figure 4: Percentage of students classified in each achievement level from each criteria in reading subtest of LEE



Similarly, those students with higher scores in reading achieved higher levels at the end of the year (See Figure 4). This pattern was true for those who attended preschool and those who did not, for boys and girls, repeaters and non-repeaters and for Mayan students and ladino students. Although the logistic regression did not identified the emergent writing as a significant predictor after controlling for others, it is possible to see



the same pattern of achievement as with the other predictors, that is those who earned more points in this test, were classified more into the satisfactory and excellent levels of achievement (See Figure 3).

Figure 3: Percentage of students classified in each achievement level from each criteria in the emergent writing subtest of LEE

What type of skills do children in each achievement level have at the end of the year?

Table 6 contains an example of an item that an average student in each achievement level in first grade could answer at the end of the year. The typical item for a student that achieves the standard at the end of first grade is similar to the typical item a typical student can answer at the beginning of second grade as it will be detailed later in this report.

Table 6: Examples of items per achievement level

Achievement level	Average theta	Item example
Unsatisfactory	-1.95	Item 8, forma A (theta= -1.24) [dibujo de niña escribiendo] a) Julio escribe b) María canta c) Julio platica d) María escribe
Must improve	-0.30	Ejemplo, forma A [figura de estrella] a) Enano b) Estrella c) Elote d) Espada
Satisfactory	0.67	Item 6, forma A (theta = 0.66) [figura de pájaro volando] a) El pájaro canta. b) El pájaro se baña. c) El pájaro vuela. d) El pájaro camina.
Excellent	2.05	Item 22, forma A (theta = 1.50) Oscar tiene las tardes libres todos los miércoles. Sus horas libres las ocupa para ser bombero voluntario. Mañana será miércoles y su mamá llegará a visitarlo. ¿Qué crees que pasará? a) Saldrán a pasear b) No lo encontrará c) Platicarán en casa d) Comerán juntos

First grade profile

Progress estimations of **First** grade students that participated in the WHIP program were done with a valid sample of 1616 students. From the sample 48% were girls and 52% were boys; 26% reported speaking a Mayan Language and the rest reported speaking Spanish. 31% did not attend preschool and the rest reported attending. Finally, 33% of the students had repeated a

grade at least once. *Students in first grade who know the alphabet principle more, and who write and read more words phonetically achieve the standard more at the end of the grade. Those students in first grade who achieved the standard can read complete sentences that match an illustration.*

Teacher profile first grade

*Progress estimations of **First** grade teachers that participated in the WHIP program were done with a valid sample of 95 teachers. From the sample 70% were female teachers and 30% were male teachers. On average, teachers in first grade had 10 years of experience; however, experience ranged from 1 to 26 years. On average, teachers had 3 students that achieved first grade standard at the end of the year, 6 that need improvement, 4 in the unsatisfactory, and 1 excellent student. On average teachers in first grade had 14 students, ranging from 3 to 21 students.*

Second Grade

As has been described before the same methodology was used for second, third and fourth grades. In the following section, results growth for second grade will be provided.

Do second grade student read more at the end of the year than they did at the beginning of the year? (Growth relative to Self)

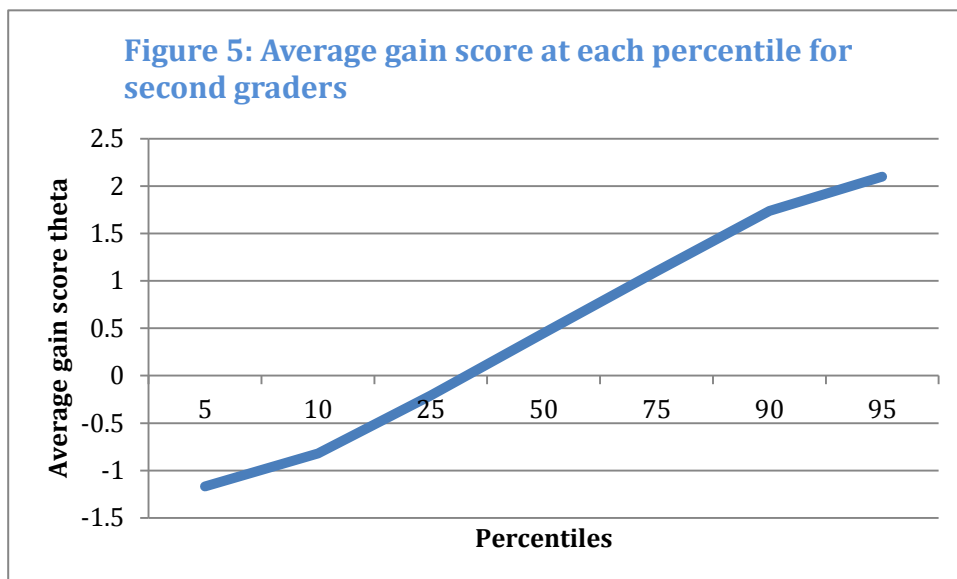
In this study, for grades second, third and fourth, gain scores were estimated by subtracting end of the year theta scores⁶ minus beginning of the year theta scores. Alternatively, end of the year raw scores minus beginning of the year raw scores were obtained. It is important to mention that the sign of the gain score has a meaning to interpret growth for each student; positive gains represent improvement and negative gains a decline. Finally, interpretation of gains in these three grades is transparent because the same test was given to students at the beginning and end of the year.

In Spanish, second grade students gained 2 points on average on the absolute scale of first grade test by the end of the year ($n= 1650$, $SD= 4.67$). The maximum gain was 22 points on this scale. On the Ability scale⁷, the gain was 0.45 points on average ($n= 1650$, $SD= 1.051$). Not all students gained ability; 66.3% of the students gained ability, and 33.7% of the students declined in ability. Out of those students who gained ability, 20% gained between 1 and 3 points, another 20% gained between 3 and 5 points, and the rest gained between 6 and 22 points.

In K'iche' students gained on average 2 points as well. However the variance of the gain was smaller than in the Spanish test ($SD=3.69$, $n= 474$). There was no gain in Mam from the beginning of the year to the end for second graders (Mean= - 0.06, Median=0, $N= 51$).

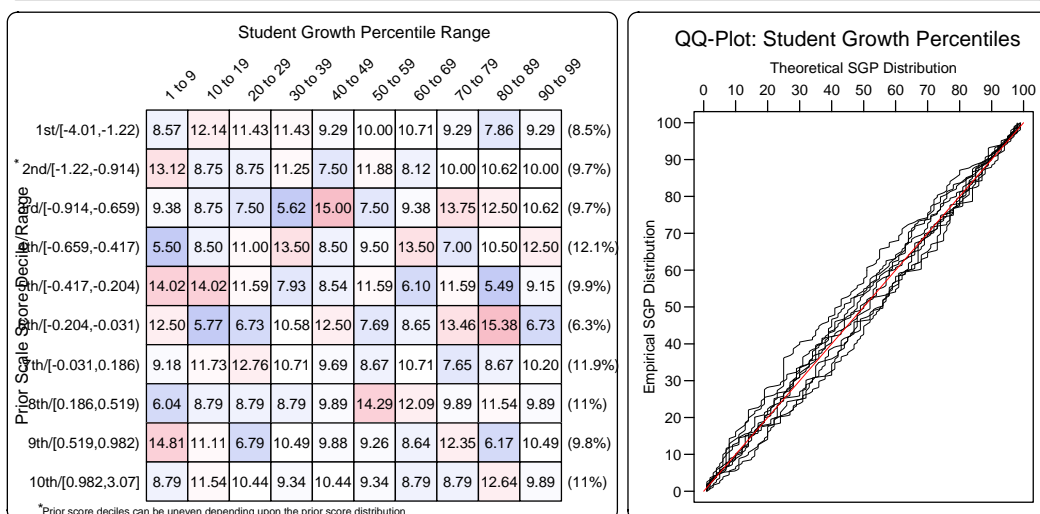
⁶ Measure of ability using Item Response Theory, specifically RASCH.

⁷ RASCH scale/theta scores



Furthermore, an average of gain scores allows group-level interpretations of progress. In second grade, group level reports were performed by sex, language, and grade repetition. The median gain for boys was 2 points and girls 3 points on the absolute scale of the first grade test. However, their mean did not differ significantly ($t = 0.470$, $df = 1648$, $p = 0.630$). Boys had an average gain theta score of 0.43 ($n = 860$) and the girls 0.46 ($n = 790$).

Student Growth Percentile Goodness-of-Fit Descriptives 2013 Reading SGP, Grade 2 (N = 1,650)



In second grade, 36% of the students had repeated a grade at least once. The average gain theta score for those who had repeated was 0.44 as opposed to 0.46 of those who had not repeated a grade. Similarly, the mean score in the absolute scale for

non-repeaters was 2.15 and 2.01 for repeaters. That is, the means did not differ for the two groups, ($t = 0.579$, $df = 1648$), $p = 0.562$.

How is second students' reading progressing in relation to their peers from the beginning of the year to the end? (Growth relative to others)

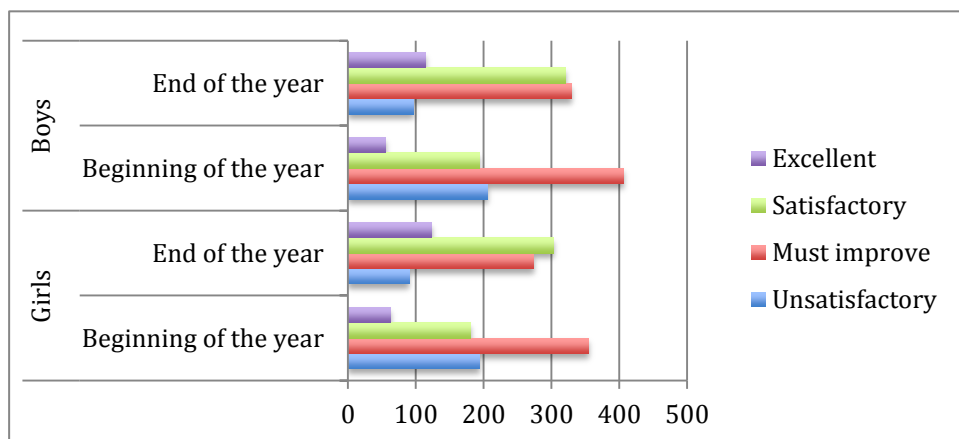
Additionally for responding if a student has made progress in reading, one might want to compare his progress to his peers; the question then is if a student is progressing from one point to the other better, worse or the same as students of the same cohort. For this purpose, the proposed model is the Student Growth Percentile (SGP) (D. Betebenner, 2009). In a few words, the SGP describes the location of a student's current score in relation to the current scores of students with similar score histories or similar scores at the beginning of the year. For example, a student earning an SGP of 80 progressed as well or better than 80% of his peers of the reference group. One shortcoming of the SGP model is that it requires large number of students for the SGP estimation to be stable; in other words if only a few students have similar scores at the beginning of the year the rank estimation becomes unstable. However, data fit analysis were done for the second grade student data, showing that sufficient fit for the sample size of second grade ($n = 1650$) (See Figure 6).

A median SGP can serve for group-level analyses. Overall, the median SGP for second graders was 50. The median SGP for girls was 51 and 48 for boys. Repeaters made less progress in the year than their peers who started at the same point at the beginning of the year. The median SGP for non-repeaters was 54 and for repeaters 43.

Is second grade students' reading closer to the established reading standard for the grade at the end of the year than their reading was at the beginning of the year? (Growth relative to standard)

Another measure of progress is to explain if a student achieved the standard or is closer to achieving it at the end of the school year than he was at the beginning of the year. In order to answer to this question, the difference between the student theta and the theta cut point for the grade will be used to estimate the proximity to achieving the standard in both points in time. In the case of second grade the interpretation is transparent as the same test was used at the beginning and end of the year. That is, the closer

Figure 7: Number of second graders in each achievement level at each time point by sex



to 0 is the difference between scores and the standard, the more progress the student has made. Average proximity to the standard for the two points in time will be used for

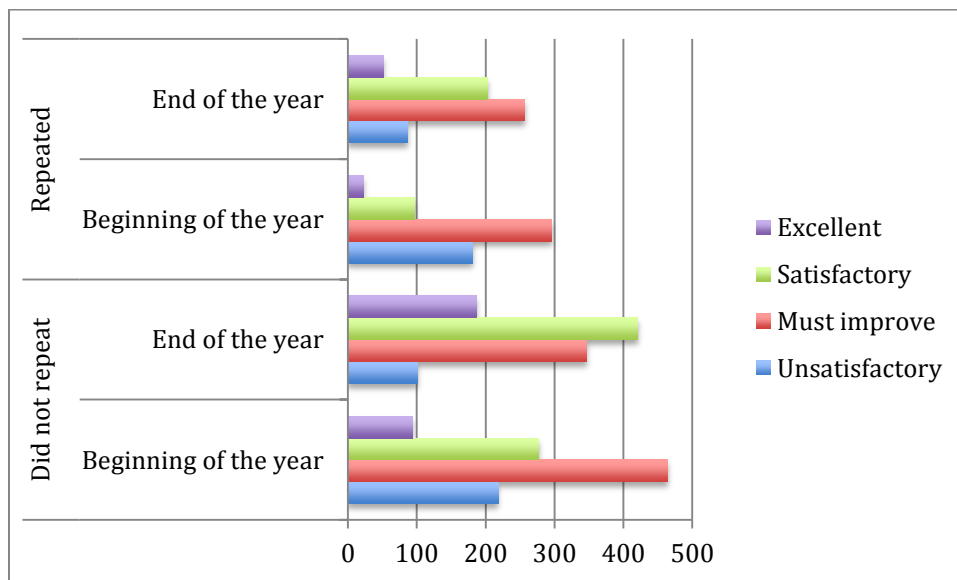
group-level interpretations. Overall students were closer to achieving the standard at the end of the year than at the beginning of the year (See Table 7).

Table 7: Average difference to first grade cut score (theta 0.201)

Time point	Average difference
Beginning of the year	-.39
End of the year	-.06

Another way to describe growth in relation to a standard is to explore the percentage of students in each achievement level at each time point. If there has been progress, students in the satisfactory level or above should have increased and the number of students in the lower levels should decrease. In second grade, this was the case. The number of students in the “satisfactory” level increased at the end of the year, and therefore, the number of students in the lower levels declined. This pattern was the same for boys and girls, and for repeaters and non-repeaters.

Figure 8: Number of repeaters and non-repeaters in each achievement level at each time point



What type of items can students in second grade answer at the beginning of the year and, what type of items can they answer at the end of the year?

At the beginning of the year, students in second grade could read, on average, items of -0.18 difficulty. An example of such items is the following:

Ejemplo, forma A

[figura de niña jugando]

a) Lucía juega.

- b) Lucía escribe.
- c) Lucía pinta.
- d) Lucía baila.

This item requires reading two words; one of them is repetitive in the distractors (name Lucía). The difficulty of this item is low or this is an easy item in the test. Furthermore, by the end of the year, students in second grade could read, on average, items of 0.26 difficulty. An example of such items is the following:

Ejemplo forma A

Al niño le _____ las mascotas. Por eso tiene un perro grande, un canario y un gato.

- a) compran
- b) agradan
- c) regalan
- d) venden

This item requires reading two related sentences and to choose the word that best fits the paragraph. The item also requires some vocabulary and knowing synonyms. This item has harder difficulty. It was interesting to notice that these two items were one page apart. This suggests that second graders made it to the third page in the test at the end of the year, which is consistent with the gain scores being small (2 points on average).

Second grade progress profile

The information above can be synthesized in the following progress profile for second graders:

*Progress estimations of **Second** grade students that participated in the WHIP program were done with a valid sample of 1650 students. From the sample 47.9% were girls and 52.1% were boys; 37.3% were ladino students, 62.3% were Mayan, 3 students were Garífuna and 3 students were Xinka. Finally, 36% of the students had repeated a grade at least once. At the beginning of the year second grade students could only answer 9 items correctly in the national reading assessment for first grade in Spanish. However, by the end of the year they answered 11 correctly on average, which represents a gain of 2 points. On average, their ability improved 0.45. An example of the type of items they can answer now filling the blank of a short paragraph with the appropriate word. This item is harder than they could do at the beginning of the year where they could only read single sentences with one subject and one verb. Unfortunately, about 33.7% declined in their ability by the end of the year. By the end of second grade, on average students achieved first grade reading standard (mean theta score= 0.26). Furthermore, the number of students in the “satisfactory” level increased at the end of the year, and therefore, the number of students in the lower levels declined. About half of the students performed better than 50% of their peers with similar ability at the beginning of second grade. The patterns in progress were about the same for girls and boys and repeaters and non-repeaters, except in the normative measure where non-repeaters progress more than their repeating peers.*

Second grade teachers' progress profile

The information above can be obtained and described per teacher. However, as a summary the following profile of second grade teachers is provided:

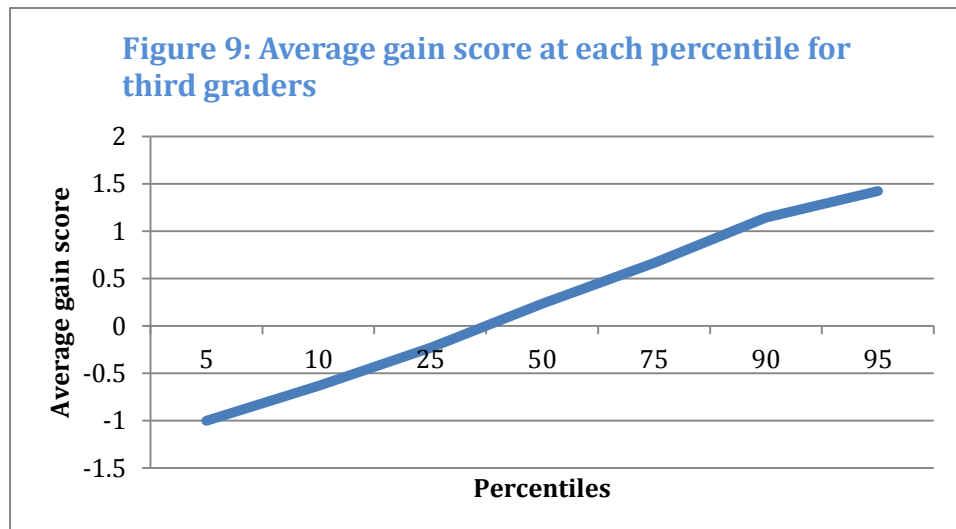
*Progress estimations of **Second** grade teachers that participated in the WHIP program were done with a valid sample of 90 teachers. From the sample 62.8% were female teachers and 37.2% were male teachers. 54.3% had less than 9 years of experience and the rest had at least 9 years of teaching experience; 9 years was the average years of experience. Students that declined were clustered in both male and female teachers, as well as in those who had more than the mean years of experience or less than the mean years of experience (9 years). Those teachers with more experience than the average, had students that progress more in relation to students that started with the same ability (median SGP= 50) than those teachers with less experience than the average (median SGP= 46).*

Third grade

As we have described previously in this study, the same methodology to describe growth in reading was applied in second, third and fourth grades. In the following section, results of third grade students will be provided by different aggregations, such as, grade, language, sex, and repetition. At the end of the section, an aggregated profile for third grade students that participated in the WHIP program is provided.

Do third grade student read more at the end of the year than they did at the beginning of the year? (Growth relative to Self)

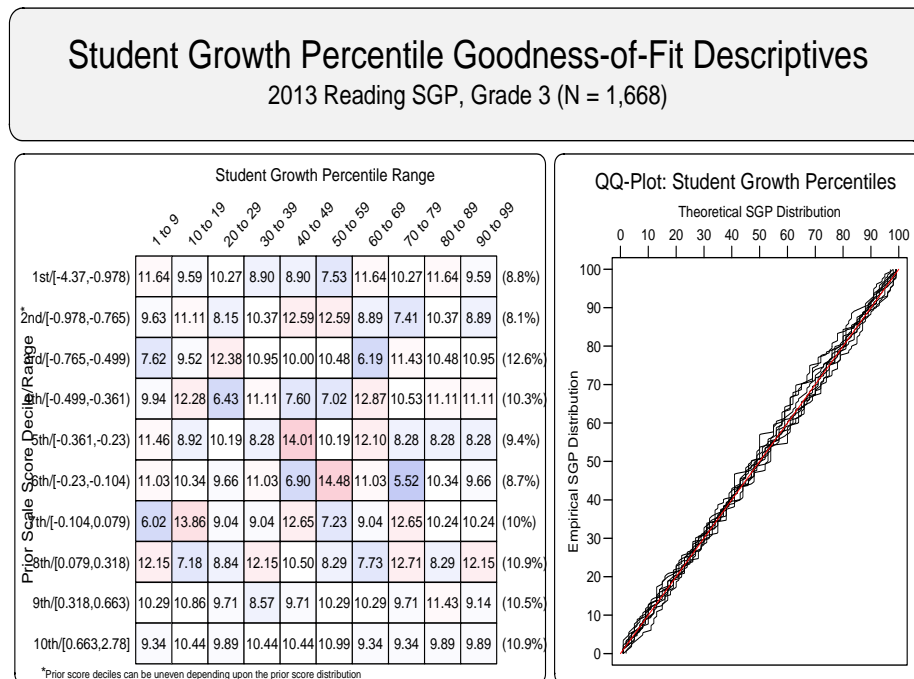
In Spanish, third grade students gained 2 points on average on the absolute scale of the test at the end of the year (n= 1668, SD= 5.25). The maximum gain was 23 points on this scale. On the Ability scale ⁸, the gain was 0.21 points on average (n= 1668, SD= 0.87). Not all students gained ability; 62.1% of the students gained ability, and 37.9% of the students declined in ability. Out of those



⁸ IRT scale/theta score

students who gained ability, 20% gained between 1 and 3 points, another 20% gained between 3 and 5 points, and the rest gained between 6 and 23 points.

In K'iche' students gain on average 1 point. However the variance of the gain was smaller than in the Spanish test ($SD=4.52$, $n= 4.73$). In Mam students also gained on average 1 point on the absolute scale ($N= 62$).



The median gain for boys was 1 point and girls 2 points on the absolute scale of the test. However, their mean did not differ significantly ($t= 1.821$, $df=1666$), $p=0.069$. Boys had an average gain theta score of 0.19 ($n= 848$) and the girls 0.22 ($n= 820$).

In third grade, 24.7% of the students had repeated a grade at least once. The average gain theta score for those who had repeated was 0.25 as opposed to 0.18 of those who had not repeated a grade. Similarly, the mean score in the absolute scale for non-repeaters was 1.72 and 1.82 for repeaters. Therefore, the means did not differ significantly for the two groups, ($t= -.328$, $df=1550$), $p=0.743$.

How is third grade students' reading progressing in relation to their peers from the beginning of the year to the end? (Growth relative to others)

Additionally for responding if a student has made progress in reading, one might want to compare his progress to his peers; the question then is if a student is progressing from one point to the other better, worse or the same as students of the same cohort. For this purpose, the proposed model is the Student Growth Percentile (SGP) (D. Betebenner, 2009). Data fit analysis were done for the third grade student data, showing sufficient fit for the sample size of third grade ($n=1668$) (See Figure 10).

Overall, the median SGP for third graders was 49. The median SGP for girls was 53 and 45 for boys. The median between repeaters and non-repeaters SGP was the same (SGP=49).

Is third grade students' reading closer to the established reading standard for the grade at the end of the year than their reading was at the beginning of the year? (Growth relative to standard)

Overall, third grade students had not achieved grade standard by the end of the year. The mean theta score was $-.01567$ and the expected theta score was at least 0.193 . Although the theta score was closer than the theta score at the beginning of the year ($\text{theta} = -.22$), students still need to improve to achieve the standard.

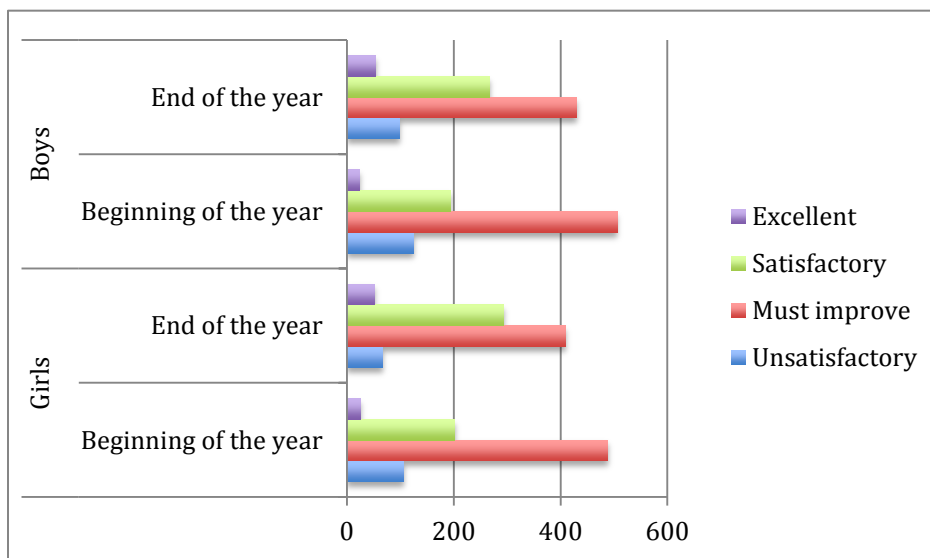


Figure 11: Number of third graders in each achievement level at each time point by repetition

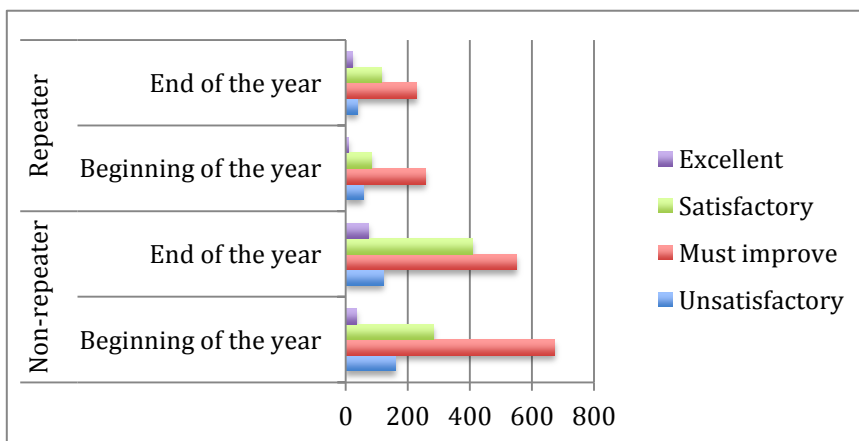


Figure 11: Number of third graders in each achievement level at each time point by repetition

Table 8: Average difference to third grade cut score (theta 0.193)

Time point	Average difference
Beginning of the year	0.41
End of the year	0.21

Another way to describe growth in relation to a standard is to explore the percentage of students in each achievement level at each time point. If there has been progress, students in the satisfactory level or above should have increased and the number of students in the lower levels should decrease. In third grade, this was the case. The number of students in the “satisfactory” level increased at the end of the year, and therefore, the number of students in the lower levels declined. This pattern was the same for boys and girls, and for repeaters and non-repeaters.

What type of items can students in third grade answer at the beginning of the year and, what type of items can they answer at the end of the year?

At the beginning of the year, students in third grade could read, on average, items of -0.22 difficulty. An example of such items is the following:

Item 14, forma A

Sergio es una persona muy amable. Todos los días ayuda a sus primos Carlos, Mario y Andrea a salir de su casa. Ellos son más pequeños y necesitan la ayuda de Sergio. ¿Quién es el personaje principal?

- a) Carlos
- b) Mario
- c) Sergio
- d) Andrea

This item required reading three sentences and to identify the main character. This item was not a hard item for third graders. By the end of the year, students in second grade could read, on average, items of -0.016 difficulty. An example of such items is the following:

Item 16, forma A

En la escuela hay actividades culturales todos los martes, jueves y viernes. Los lunes y miércoles se practica deporte. El lunes las niñas juegan básquet y los niños atletismo. El miércoles los niños juegan básquet mientras las niñas hacen atletismo. ¿Qué día juegan básquet las niñas?

- e) Lunes
- f) Martes
- g) Miércoles
- h) Jueves

This item is harder because it requires reading more sentences in the paragraph and to remember the details of it. It is interesting to notice that this item is only two items apart from the item students in third grade could read at the beginning of the year, which is consistent with the gain score.

Third grade progress profile

The information above can be synthesized in the following progress profile for third graders:

*Progress estimations of **Third** grade students that participated in the WHIP program were done with a valid sample of 1668 students. From the sample 45.9% were girls and 47.5% were boys; 42.1% were ladino students, 50.5% were Mayan, 7 students were Garífuna and 5 students were Xinka. Finally, 24.7% of the students had repeated a grade at least once. At the beginning of the year third grade students could only answer 12 items correctly in the national reading assessment for third grade in Spanish. However, by the end of the year they answered 14 correctly on average, which represents a gain of 2 points. On average, their ability improved 0.21. An example of the type of items they can answer now is finding the details in a five-sentence paragraph. This item is harder than they could do at the beginning of the year where they could only find the main character in a four-sentence paragraph. Unfortunately, about 37.9% declined in their ability by the end of the year. By the end of third grade, on average students still had not achieved third grade reading standard but they are closer to achieving it than they were at the beginning of third grade. Furthermore, the number of students in the “satisfactory” level increased at the end of the year, and therefore, the number of students in the lower levels declined. Girls progressed more in relation to their peers that started with the same ability.*

Third grade teachers' progress profile

The following is an illustration of a teacher progress profile that could be created based on the information obtained from the different growth models described in the previous section and the data available for the WHIP program.

- *Progress estimations of **Third** grade teachers that participated in the WHIP program were done with a valid sample of 94 teachers. From the sample 57.4% were female teachers and 42.6% were male teachers. 50% had less than 10 years of experience and the rest had at least 10 years of teaching experience; 10 years was the average experience. Students that declined were clustered in both male and female teachers, as well as in those who had more than the mean years of experience or less than the mean years of experience (9 years). Those teachers with more experience than the average, had students that progress more in relation to students that started with the same ability (median SGP= 50) than those teachers with less experience than the average (median SGP= 46).*

Fourth Grade

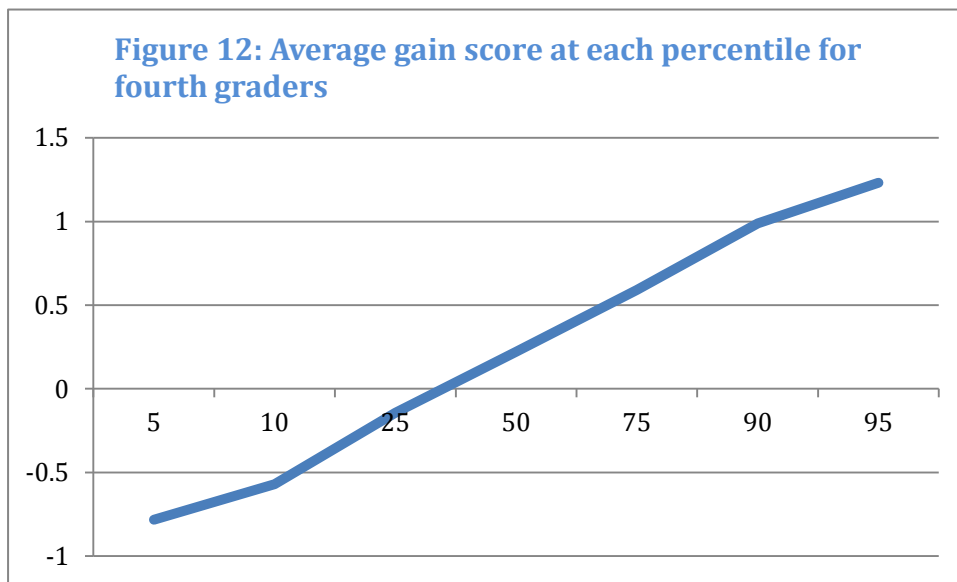
As we have described previously in this study, the same methodology to describe growth in reading was applied in second, third and fourth grades. In the following section, results of fourth grade students will be provided by different aggregations, such as, grade, language, sex, and repetition. At the end of the section, an aggregated profile for fourth grade students that participated in the WHIP program is provided.

Do fourth grade student read more at the end of the year than they did at the beginning of the year? (Growth relative to Self)

In Spanish, fourth grade students gained 2 points on average on the absolute scale of the test at the end of the year ($n=1612$, $SD=4.89$). The maximum gain was 20 points on this scale. On the Ability scale, the gain was 0.21 points on average ($n=1612$, $SD=0.65$). Not all students gained ability; 67.2% of the students gained ability, and 32.8% of the students declined in ability. Out of those students who gained ability, most of them gained between 1 and 4 points.

In K'iche' the collection was done only at the end of the year; therefore, calculations of growth are not possible. There was no gain in Mam from the beginning of the year to the end for fourth graders

(Mean= 0.7, Median=0, N= 20).



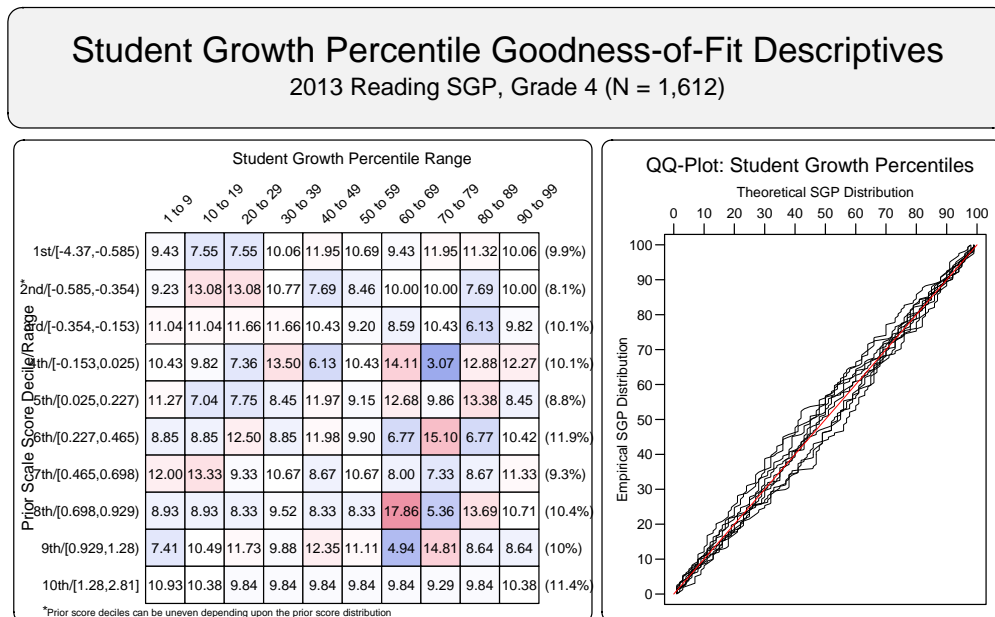
The median gain for boys was 1 point and girls 2 points on the absolute scale of the test. However, their mean did not differ significantly ($t=0.151$, $df=1603$, $p=0.88$). Boys had an average gain theta score of 0.21 ($n=816$) and the girls 0.22 ($n=789$).

In fourth grade, 21.4% of the students had repeated a grade at least once. The average gain theta score for those who had repeated was 0.22 as opposed to 0.20 of those who had not repeated a grade. Similarly, the mean score in the absolute scale for non-repeaters was 1.66 and 1.85 for repeaters. Therefore, the means did not differ for the two groups, ($t=-0.599$, $df=1496$), $p=0.549$.

How is fourth grade students' reading progressing in relation to their peers from the beginning of the year to the end? (Growth relative to others)

Additionally for responding if a student has made progress in reading, one might want to compare his progress to his peers; the question then is if a student is progressing from one point to the other better, worse or the same as students of the same cohort. For this purpose, the proposed model is the Student Growth Percentile (SGP) (D. Betebenner, 2009). Data fit analysis were done for the fourth grade student data, showing that sufficient fit for the sample size of second grade (n=1650) (See Figure 13).

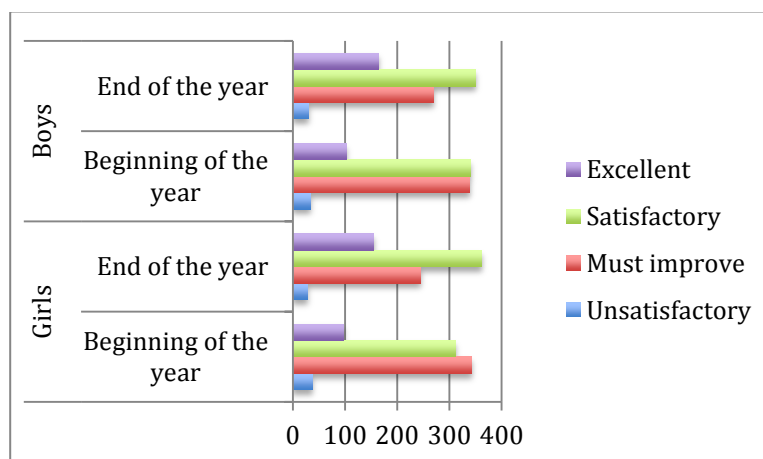
A median SGP can serve for group-level analyses. Overall, the median SGP for fourth graders was 49. The median SGP for girls was 52 and 47 for boys. The median between repeaters was 49 and 50 for non-repeaters.



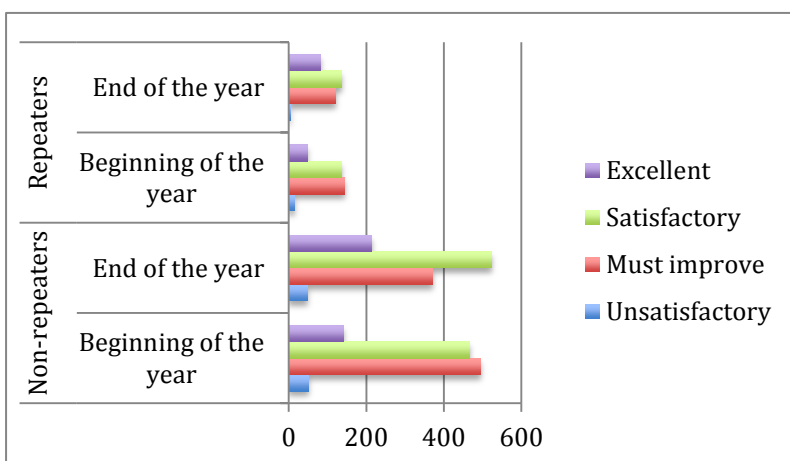
Is fourth grade students' reading closer to the established reading standard for the grade at the end of the year than their reading was at the beginning of the year? (Growth relative to standard)

On average, fourth grade students had achieved grade three standard since the beginning of the year ($\theta=0.30$); however, did not made it to excel it at the end of the year ($\theta=0.51$). The expected satisfactory θ score for third grade is 0.193, and the excellent θ score is 1.24.

Figure 15: Number of fourth graders at each achievement level and time point by sex



Another way to describe growth in relation to a standard is to explore the percentage of students in each achievement level at each time point. If there has been progress, students in the satisfactory level or above should have increase and the



number of students in the lower levels should decrease. In fourth grade, this was the case. The number of students in the “satisfactory” level increased at the end of the year, and therefore, the number of students in the lower levels declined. This pattern was the same for boys and girls, and for repeaters and non-repeaters.

Figure 14: Number of fourth graders at each achievement level and time point by repetition

What type of items can students in fourth grade answer at the beginning of the year and, what type of items can they answer at the end of the year?

At the beginning of the year, students in fourth grade could read, on average, items of 0.30 difficulty. An example of such items is the following:

Item 15 forma A

Julio y Ana son buenos amigos y les gustan los animales. Julio tiene un pato blanco y un perro peludo. Ana tiene un gato peludo y una Tortuga verde. ¿Qué tiene Ana?

- a) Un gato blanco y un perro peludo.
- b) Un gato peludo y una tortuga verde.
- c) Un pato blanco y tortuga verde.
- d) Un gato peludo y una tortuga blanca.

This item required reading four sentences and to remember details of the paragraph. It is interesting that this was the ending point of third graders. By the end of the year, students in fourth grade could read, on average, items of 0.51 difficulty. An example of such items is the following:

Item 26, forma A

El Niño es un fenómeno natural causado por las temperaturas altas en el Pacífico y las aguas de la costa de Centro América. Afecta de alguna manera a todas las regiones del planeta y nota más durante el invierno. ¿Cuál es el propósito del autor o autora del texto?

- a) Informar
- b) Criticar
- c) Cuestionar
- d) Convencer

This item is much harder because it requires reading more sentences in the paragraph and to infer the purpose of the author.

Fourth grade progress profile

The information above can be synthesized in the following progress profile for fourth graders:

*Progress estimations of **Fourth** grade students that participated in the WHIP program were done with a valid sample of 1605 students. From the sample 49% were girls and 51% were boys; 38.6% were ladino students, 54.5% were Mayan, 3 students were Garífuna. Finally, 24.7% of the students had repeated a grade at least once. At the beginning of the year fourth grade students could only answer 12 items correctly in the national reading assessment for third grade in Spanish. However, by the end of the year they answered 14 correctly on average, which represents a gain of 2 points. On average, their ability improved 0.21. An example of the type of items they can answer now is inferring the purpose of the author in a four-sentence story. This item is harder than they could do at the beginning of the year where they could only identify details of a short paragraph. Unfortunately, about 37.9% declined in their ability by the end of the year. Since the beginning of fourth grade, on average students had achieved third grade reading standard but they did not make it on average to excel third grade standard at the end of fourth grade. Girls progressed more in relation to their peers that started with the same ability.*

Fourth grade teachers' progress profile

The following is an illustration of a teacher progress profile that could be created based on the information obtained from the different growth models described in the previous section and the data available for the WHIP program.

- Progress estimations of **Fourth** grade teachers that participated in the WHIP program were done with a valid sample of 94 teachers. From the sample 51.1% were female teachers and 48.9% were male teachers. 45.5% had at least 11 years of experience; 11 years was the average experience. Students that declined were clustered in both male and female teachers, as well as in those who had more than the mean years of experience or less than the mean years of experience (11 years). Those teachers with more experience than the average, had students that progress more in relation to students that started with the same ability (median SGP= 53) than those teachers with less experience than the average (median SGP= 49).

V. CONCLUSIONS

All the models used in this study coincide in results to describe improvement in reading. This study allowed describing improvement under different perspectives of growth. Table 9 summarizes the different estimations done for each of the approaches to growth in each grade.

Table 9: Summary statistics of growth under each approach

		GRADE AVERAGE	BOYS	GIRLS	REPEATERS	NON REPEATERS
Growth relative to self: GAIN SCORE (THETA SCALE)**	SECOND	0.45	0.43	0.46	0.44	0.46
	THIRD	0.21	0.19	0.22	0.25	0.18
	FOURTH	0.21	0.21	0.22	0.22	0.2
Growth relative to self: GAIN SCORE (ABSOLUTE SCALE)**	SECOND	2	2	3	2	2
	THIRD	2	1	2	2	2
	FOURTH	2	1	2	2	2
Growth relative to others: SGP*	SECOND	50	48	51	43	54
	THIRD	49	45	53	49	49
	FOURTH	49	47	52	49	50
Growth relative to a standard: Percentage of standard achievers, at the end of the year	SECOND	52%	50%	54%	42%	58%
	THIRD	40%	38%	42%	37%	42%
	FOURTH	64%	63%	65%	64%	64%

Children that participated in the WHIP program improved in reading; however the improvement is low. Overall, children in second, third and fourth grades improved during the school year in the five departments where the Western Highlands Integrated Program – WHIP. Gains ranged from 1 to 3 points on the absolute scales, and between 0.18 and 0.46 on the theta scale.

Students improved more in second grade than in the upper grades. Students in second grade gained between 0.43 and 0.46 points in the theta scale; however, in the upper grades (third and fourth) the gained was between 0.2 and 0.22 points.

Children that participated in the WHIP program improved less in reading in the Mayan languages than they do in Spanish. Students showed a little improvement in K'iche'. And, with the exception of grade three, there was no improvement in second and fourth grades (See Table 10).

Table 10: Absolute gain in each language by grade

	Second	Third	Fourth
Gain in the absolute scale K'iche'	2	1	No data available
Gain in the absolute scale Mam	0	1	0

A high percentage of students that participated in the WHIP program declined in ability. Most students gained ability; however, in second grade 33.7% of students declined. In third the percentage was 37.9 and in fourth grade 32.8.

Girls that participated in the WHIP program progressed more in reading than boys. In all the models used to described progress in this study, girls had higher estimations than boys (See Table 9).

Repeaters improved less than their non-repeaters peers. Non-repeaters and repeaters had about the same gains during the year, except in third grade, where repeaters gained more. However, in relation to their peers, repeaters consistently progressed less than their non-repeater peers. Furthermore, non-repeaters achieved the standard more than repeaters.

Students that know the alphabetic principle, read words and write some words, are more likely to achieve first grade standard at the end of the year. The criteria to classify students in LEE are positively related to the level of achievement at the end of first grade. First grades with higher scores in the Emergent reading assessment LEE at the beginning of the year are more likely to achieve first grade standard at the end of the year. Similarly, those students with lower scores in LEE are more likely to be classified in the “unsatisfactory” and “needs improvement” levels of achievement.

Students that participated in the WHIP program are a year behind in terms of achieving the standard. Students in first grade on average do not achieve first grade standard at the end of first grade or second grade. However, they are closer to achieving it at the end of second grade. Third graders do not achieve third grade standard at the

end of the year but they do in fourth grade. Although, on average fourth grade students achieve third grade standard, they do not excel it at the end of the year.

It is possible to build a learning continuum from the average items that students read at each grade and time point. In terms of the skills that students have, it was possible to build a progress latter that illustrates the skills that students have at each grade from the typical items of entry and exit of each grade. Figure illustrates such progress in terms of skills students have.

Table 11: Illustration of progress in skills

End of the year →	Fourth grade		Third grade		Second grade		First grade		LEE	
	Infers author's purpose in a paragraph.		Remembers details of a three-sentence paragraph		Finds the appropriate word to complete a paragraph.		Read two words to match a picture (one word is a name).		Reads words	
	Remembers details of a four-sentence paragraph.		Identifies the main character of a three-sentence paragraph		Read two words to match a picture (one word is a name).		Read one word that matches a picture.		Knows alphabet	
Beginning of the year →										

VI. IMPLICATIONS

Measuring learning implies longitudinal studies. As with this study, measuring learning implies collecting achievement information at least twice for each student. This methodology implies complex logistics; however, conclusions about learning are more robust.

Interventions focused to teachers should be based on learning and progress information. In this study it was possible to create profiles of teachers focused on their students' learning or progress. This will allow creating interventions that are focused to teachers that need more help in bringing their students to a higher achievement level.

Progress information should be based on different models that coincide in results. Descriptions or conclusions about progress or learning should be based in different models that coincide in results. This methodology allows stronger conclusions at the individual and group levels.

An effort to communicate progress/growth in reading to teachers should be made. Illustrations of a learning continuum in each grade should be provided to teachers that orient their teaching practice.

VII. REFERENCES

- Betebenner, D. (2009). Norm-and Criterion-Referenced Student Growth. *Educational Measurement: Issues and Practice*, 28(4), 42–51.
- Betebenner, D. W. (2008). A primer on student growth percentiles. *Dover, NH: National Center for the Improvement of Educational Assessment*. Retrieved February, 18, 2011.
- DePascale, C. A. (2006). Measuring growth with the MCAS tests: A consideration of vertical scales and standards. *Dover, NH: National Center for Improvement in Educational Assessment*. Retrieved on March, 29, 2012.
- Ho, A. D., & Castellano, K. E. (2013). A Practitioner's Guide to Growth Models. Presented at the Council of Chief State School Officers, Cambridge, MA: Council of Chief State School Officers. Retrieved from <http://scholar.harvard.edu/andrewho/publications/practitioners-guide-growth-models>

USAID/Reforma Educativa en el Aula

Avenida La Reforma 6-64 zona 9
Plaza Corpotativa Reforma, Torre II
Nivel 9, Oficina 901
Guatemala, C.A.

Tel: (502) 2390-6700

website:reaula.org

Correo electrónico:reaula@juarezassociates.com